

**UNITED STATES DISTRICT COURT
DISTRICT OF MINNESOTA**

In re Bair Hugger Forced Air Warming
Products Liability Litigation

MDL No. 15-2666 (JNE/FLN)

This Document Relates To:
ALL CASES

**DEFENDANTS' OPPOSITION TO PLAINTIFFS' MOTIONS TO EXCLUDE
TESTIMONY OF THEODORE HOLFORD AND JONATHAN BORAK**

TABLE OF CONTENTS

INTRODUCTION	1
BACKGROUND: THE MCGOVERN STUDY	4
ARGUMENT	6
I. BY IDENTIFYING TABULATION ERRORS AND EVALUATING THE IMPACT OF DISCLOSED CONFOUNDERS, HOLFORD AND BORAK DEMONSTRATE THAT THE MCGOVERN STUDY IS NOT RELIABLE	6
A. Holford Used Reliable Methods to Perform a Biostatistical Analysis of the McGovern Data and Identify Flaws in the Study	6
1. Plaintiffs’ theory that the data Holford analyzed are not the “real” McGovern data is contradicted by the evidence and Plaintiffs’ own in- court representations	7
2. Holford identified a serious tabulation error in McGovern’s underlying data and reanalyzed the data to show the impact.	11
3. McGovern’s co-author Mike Reed recalled the tabulation error slightly differently, and Holford also accounted for the possibility that Reed’s recollection was accurate	12
4. Holford and Borak demonstrate the fatal impact of disclosed confounders on McGovern’s reliability	15
i. Confounder 1: Rivaroxaban	16
ii. Confounder 2: Antibiotic Prophylaxis	20
5. Holford’s analysis demonstrates that the chosen start date affects McGovern’s results and thus its reliability.	25
II. BORAK APPLIED A VALID METHODOLOGY TO IDENTIFY ADDITIONAL UNDISCLOSED CONFOUNDERS IN MCGOVERN	33
A. Massive Infection Reduction Measures Were Undertaken at the Hospital Studied by McGovern—Measures that Disproportionately Benefited the Non-Bair Hugger Patients	33

B.	Plaintiffs Fail to Rebut Borak’s Conclusion that These Infection Control Measures Undermine McGovern’s Validity.....	35
C.	The Remainder of Plaintiffs’ Attack on Borak Consists of Misrepresentations of His Report and His Methods	36
III.	PLAINTIFFS’ ARGUMENT THAT MCGOVERN IS RELIABLE <i>PER SE</i> IS CONTRARY TO CASE LAW	36
A.	Plaintiffs’ Cite to the <i>PPA</i> Case Only Underscores Why McGovern Is Not Reliable and the Importance of Borak and Holford’s Testimony.....	37
B.	Unlike the HSP Discussed in <i>PPA</i> , McGovern’s Severe Flaws Compromise Its Findings	39
IV.	HOLFORD AND BORAK MEET THE REQUIREMENTS OF <i>FRYE-MACK</i>	43
	CONCLUSION	43
	ADDENDUM	46

INTRODUCTION

The Court should deny Plaintiffs' motions to exclude Defendants' experts Theodore Holford, Ph.D., and Jonathan Borak, M.D. Holford and Borak give qualified, reliable, and relevant opinions establishing the lack of reliability of the key study relied upon by Plaintiffs and their experts: the 2011 McGovern study.¹

McGovern is the foundation of the general causation opinions of Plaintiffs' medical experts, Drs. Samet, Jarvis, and Stonnington. It is the only study that purports to show an association between Bair Hugger use and surgical site infections in real-world procedures. Despite the study's concession that "[t]his study does not establish a causal basis" for the association, despite the study's origins in Dr. Scott Augustine's "publication factory," and even though it is a comparison of the Bair Hugger system and Augustine's competitor product rather than against background risk, Plaintiffs and their experts opine that it does, in fact, establish causation. Holford and Borak address McGovern's reliability and rebut Plaintiffs' medical experts' opinions that depend upon it. *See Aviva Sports, Inc. v. Fingerhut Direct Mktg., Inc.*, 829 F. Supp. 2d 802, 835 (D. Minn. 2011) (Ericksen, J.) ("It is the proper role of rebuttal experts to critique plaintiffs' expert's methodologies and point out potential flaws in the plaintiff's experts' reports.").

Holford is the Susan Dwight Bliss Professor of Epidemiology and Public Health (Biostatistics) at the Yale University School of Public Health, where he has taught for

¹ DX5, McGovern et al., "An Investigation of Theatre Ventilation, Patient Warming and Joint Replacement Infection in Orthopedics," 93B:11 *J. Bone & Joint Surg.* 1537 (2011). Cites to "DX" are to the exhibits to the Declaration of Corey L. Gordon, filed concurrently.

more than forty years. He is a biostatistician who develops and applies statistical methods in public health and health research. Holford audited McGovern's results, identified several flaws in the study, including tabulation errors, and reanalyzed the study's data to account for those flaws. He concludes that, when one controls for just the two disclosed confounding factors, there is no difference in infection rates between the Bair Hugger system and Scott Augustine's HotDog product. The "association" disappears.

Borak is a Clinical Professor of Epidemiology and Public Health and Clinical Professor of Medicine at Yale, where he has taught for more than thirty years. He is a medical doctor who teaches interpretation of epidemiological data and inferences of causation. He relies on Holford's reanalysis and employs his own expertise to expand on Holford's discussion and address additional confounders that were not disclosed in the McGovern study. *See Reference Manual on Scientific Evidence* 215 (3d ed. 2011) (noting that statisticians typically are disclosed alongside a subject-matter expert who applies the statistical analysis). This includes, most significantly, a massive infection control initiative undertaken at the hospital studied by McGovern—an initiative that disproportionately benefited the non-Bair-Hugger patients. Finding McGovern to be fundamentally flawed on these bases, Borak concludes that the opinions of Plaintiffs' medical causation experts are not scientifically supported. *See In re Mirena IUD Prods. Liab. Litig.*, 169 F. Supp. 3d 396, 418–19 (S.D.N.Y. 2016) (“[G]iven that Defendants’ experts are attempting to prove a negative . . . pointing to the absence of convincing studies or the weaknesses of studies on which Plaintiffs rely, and evaluating them in light

of their own . . . experience, training and research, is in these circumstances a logical and valid approach.”).

In short, Holford and Borak applied the same methods that they teach to their students at Yale to dissect McGovern and rebut Plaintiffs’ experts who rely on McGovern.

Most of Plaintiffs’ attack on Holford and Borak is premised on misrepresenting their methods and the scientific literature. Their flailing underscores the importance of Holford and Borak’s opinions. Holford and Borak lay out the truth about McGovern’s fatal flaws. Not only are their opinions methodologically sound (which is all the Court needs to find to deny Plaintiffs’ challenge), they also prove the utter *lack* of reliability of McGovern and the opinions of Plaintiffs’ experts who rely upon it. Their opinions further confirm what fact discovery had already uncovered: the data and analyses underlying the McGovern study were manipulated to accomplish the aims of Dr. Augustine, who employed Mark Albrecht, the statistician who performed the data analysis.

Plaintiffs also argue that, even if Holford and Borak’s dissection of McGovern is correct, McGovern is *per se* reliable because it was peer-reviewed. Case law is to the contrary. Courts have concluded that even peer-reviewed studies, when based upon faulty data or analysis as McGovern is here, are not reliable. *See, e.g., In re Viagra Prods. Liab. Litig.*, 658 F. Supp. 2d 936, 945, 950 (D. Minn. 2009) (“*Viagra II*”). (“Peer review and publication mean little if a study is not based on accurate underlying data.”). That is what Judge Magnuson concluded in the Viagra MDL as a prelude to excluding

the plaintiffs' general causation expert and granting summary judgment to the defendant. *See id.*

For these reasons, and the additional reasons discussed below, Holford and Borak's opinions and testimony are admissible under both Fed. R. Evid. 702 and Minn. R. Evid. 702. Plaintiffs' motions to exclude Holford and Borak should be denied.

BACKGROUND: THE MCGOVERN STUDY

Holford's biostatistical analysis, and Borak's extension of that analysis, are highly relevant because Plaintiffs' case depends on McGovern. Without it, their causation evidence is purely theoretical; they lack any evidence that the Bair Hugger system increases the incidence of surgical site infections in the real world. An explanation of the study's background, development, and methodology helps make this clear.

The McGovern study monitored infection rates at Wansbeck Hospital in Northumbria, England, during a period when the Bair Hugger system was in use versus a period when Augustine's HotDog was in use. It was an observational study, meaning that it was not blinded and controlled like a clinical study.² DX13, Albrecht Dep. 132:3–134:10. *See In re Prempro Prods. Liab. Litig.*, 738 F. Supp. 2d 887, 891 (E.D. Ark. 2010) (“Because this procedure lacks controls, observational studies are more susceptible to bias and other confounding factors, and so are less reliable than clinical studies, which

² This type of observational study is known as an “interrupted time series, or “ITS.” DX7, Samet Dep. I at 93:13-14. In an ITS, data are collected before and after an intervention to detect whether the intervention has had an effect significantly greater than the secular trend. *See* DX8, Ramsay et al., “Interrupted Time Series Designs in Health Technology Assessment: Lessons from Two Systematic Reviews of Behavior Change Strategies,” 19(4) *Int. J. Tech. Assess. Health Care* 613 (2003) (concluding that “ITS designs are often analyzed inappropriately, under powered, and poorly reported”).

are often referred to as the ‘gold standard.’”).

McGovern co-author Mike Reed testified that the study’s initial objective was different. It consisted of an experiment to determine whether the Bair Hugger system had a greater effect on operating room airflow than Augustine’s HotDog. DX9, Reed Dep. 25:14-25. After this experiment, the authors decided to look at infection rates in the Bair Hugger and HotDog periods. This comparison was, in Reed’s words, “more opportunistic.” *Id.* at 25:17-21.

For the Bair Hugger period, the authors collected data on 1,066 procedures over 20 months. They collected data on just 371 procedures over seven months for the HotDog period. They then conducted statistical runs on the data until they had, in the words of the paper’s statistician (and Augustine employee) Mark Albrecht, “barely made it [to statistical significance] with the new sample numbers.” DX11, 1/11/2011 Email from Albrecht to Reed at 1. Having come up with just the right combination of time periods that achieved the Holy Grail of a “significant” difference, Albrecht updated the draft to “reflect the new infection numbers.” *Id.*

At the behest of peer reviewers, the authors were forced to provide details about two potential confounding factors to which they had vaguely alluded in earlier drafts. DX12, Response to Reviewer Comments at 2560. The authors described two changes during the study period that could have explained the apparent reduction in infections: (i) a change in the antibiotic regimen, and (ii) a change in the drugs given to prevent clotting (thromboprophylaxis). McGovern 1543. The authors mentioned these two changes “for example” to support their statement that “the data are observational and may be

confounded by other infection control measures instituted by the hospital.” *Id.* No other “examples” were mentioned.

The published study did not attempt to control for either confounder or the differences in patient characteristics. *Id.* (“[W]e were unable to consider all factors that have been associated with SSI, as the details of blood transfusion, obesity, incontinence and fitness for surgery, which have been identified elsewhere as important predictors for deep infection, were not sufficiently detailed in the medical record.”). Instead, the authors simply compared the average joint infection rate during the 20-month Bair Hugger period to the average infection rate during the later seven-month HotDog period. *Id.* The paper reported 32 infections during the Bair Hugger period and only three infections during the HotDog period, resulting in a calculated odds ratio of 3.8 (Bair Hugger infection rate divided by HotDog infection rate). *Id.* 1542. It is this 3.8 odds ratio to which Plaintiffs cling, and upon which their experts’ opinions—and their entire case—depends.

ARGUMENT

I. BY IDENTIFYING TABULATION ERRORS AND EVALUATING THE IMPACT OF DISCLOSED CONFOUNDERS, HOLFORD AND BORAK DEMONSTRATE THAT THE MCGOVERN STUDY IS NOT RELIABLE.

A. Holford Used Reliable Methods to Perform a Biostatistical Analysis of the McGovern Data and Identify Flaws in the Study.

Based upon his biostatistical analysis of the data underlying the McGovern study, Holford concludes (i) that significant tabulation errors were made that resulted in an incorrect result being published; and (ii) that when just two of the disclosed confounding

variables are accounted for, there is no difference in the infection rates between the Bair Hugger period and the HotDog period. Borak relies on Holford's analysis to support his rebuttal to the general causation opinions of Plaintiffs' medical experts.

1. Plaintiffs' theory that the data Holford analyzed are not the "final" McGovern data is contradicted by the evidence and Plaintiffs' own in-court representations.

Plaintiffs' first attack is to deny that the data Holford analyzed are, in fact, the actual data underlying the McGovern study. Pl. Holford Mem. 9. They speculate that the data Holford reviewed were not the "final" McGovern data. *Id.* Plaintiffs' attack is contradicted by the discovery record and their own recent representations to the Court.

From the first days of discovery, Defendants aggressively pursued every potential source of information on the McGovern study. Albrecht, who performed the data analysis, was an obvious first stop. Albrecht testified that he no longer had the data himself but that Augustine would. Albrecht Dep. 150:4–151:3. Defendants subpoenaed Augustine for documents relating to the McGovern study, including the underlying data, and the data analyzed by Holford were the only data Augustine produced in response. Because English law precluded compelled document discovery from the U.K.-based study authors, Defendants had no means to require anyone else to produce the data. Thus, the dataset produced by Augustine was the only complete set of McGovern infection data produced in this litigation. DX14, McGovern Dataset Produced by Augustine.

All the evidence supports the conclusion that the data Augustine produced in response to Defendants' subpoena are the "real" McGovern data:

First, except for the tabulation error identified by Holford, the data produced by Augustine correspond exactly to the representations about the data in the published McGovern study. *See* McGovern 1541.

Second, Reed testified that the data produced by Augustine were almost certainly the "real" data (that is, the dataset provided by Reed to Albrecht, as discussed below):

Q. 788 through 1081. Does that look familiar to you?³

A. Yes.

Q. Is that the form of the data on infections that you would have pulled and provided to your co-authors?

A. Yes.

Q. Who did the actual data analysis?

A. For this paper, Mark Albrecht.

Q. So were these data, pages 788 through 1081, provided by you to Mr. Albrecht?

MR. ASAAD: Objection, lack of foundation.

THE EXAMINER: You may answer.

A. I expect so. I don't remember that, but I imagine I did. There was nothing on here that would -- you know, there is no data governance issues with this. So I think, I am almost certain I would have provided it.

...

³ Albrecht Exhibit 10 was marked as part of a Reed group exhibit, Reed Ex. 3 at 788–1081. Thus, Reed's quoted testimony relates to Albrecht Exhibit 10.

MR. ASAAD:

I would just like a clarification for my objection. I am uncertain whether or not this witness is saying that this is exactly what he gave or used, or whether he says it looks like it, but he is not clear. I just want a clarification.

THE EXAMINER: Which is it, Mr. Reed?

A. In all honesty, it looks like it. I don't know if it is what I gave. But I don't know where he would have gotten it, if it wasn't from me.

MR. ASAAD: Do you know whether or not it's accurate?

A. I don't know.

Reed Dep. 52:22–54:4; *see also generally id.* 54:19–64:7 (where Reed demonstrates his familiarity with the data set by explaining the meaning of each of the abbreviations and data categories on the Excel spreadsheet).

Third, when reviewing the data at his deposition, Albrecht deduced that they were probably the updated data that had been the subject of Reed's private communication with Albrecht—the data that supported a lower odds ratio than the published study. Albrecht Dep. 163:8-24.

Fourth, though Defendants could not compel discovery from him under English law, Dr. McGovern did voluntarily produce all documents in his files related to his Bair Hugger research, including a spreadsheet containing a subset of data for procedures that resulted in infections at the study hospital (Wansbeck General). DX15, Wansbeck General Hospital Data Subset. The entries on McGovern's spreadsheet correspond exactly with the data produced by Augustine on such parameters as date of surgery, age

and sex of patient, type of procedure, date of infection diagnosis, description of bacteria, etc. Both data sets reflect a total of 31 procedures during the 20-month Bair Hugger period, and four infections during the seven-month HotDog period, that resulted in infection. The only difference is in the classification of warming devices: the dataset produced by Augustine does not indicate whether the procedure used the Bair Hugger system (forced air warming, or “FAW”) or the HotDog (conductive warming blanket, or “CWB”). In contrast, fields were added to McGovern’s spreadsheet that purported to demonstrate whether the procedure involved “FAW” or “CWB.” On questioning by Plaintiffs’ counsel, McGovern agreed that the spreadsheet he produced “match[es] the data presented in the published study.” DX16, McGovern Dep. II 371:3-6.

Finally, Plaintiffs had previously *conceded* that the data Holford looked at are the “real” McGovern data. At argument on Plaintiffs’ Motion for Leave to Amend to Assert Punitive Damages, counsel for Plaintiffs represented to the court that their epidemiology expert, Samet, had looked at “all the evidence, *all the underlying raw data*,” and concluded that McGovern is “absolutely a valid study.” 5/18/2017 Hrg. Tr. 32:2-4 (emphasis added).⁴ Plaintiffs’ newfound belief that the data analyzed by Holford are not the “real” data is not credible when Plaintiffs took no effort to obtain the supposed “real” data from any source. Certain of Plaintiffs’ counsel (Kennedy Hodges) have represented and worked with Augustine since 2009. ECF No. 231 at 5. One would think that, if there

⁴ Contrary to Plaintiffs’ representation, Samet later testified that he had not seen the McGovern data before rendering his opinion and had not reanalyzed the data. Samet Dep. 44:8–45:19.

were other, “real” data out there, Plaintiffs would have a means (and the incentive) to get them.

2. Holford identified a serious tabulation error in McGovern’s underlying data and reanalyzed the data to show the impact.

Plaintiffs also argue that Holford “remixed” and manipulated the data to reach his conclusions. Pl. Holford Mem. at 8–9, 17. He did nothing of the sort, and Plaintiffs’ contention is based entirely upon mischaracterizing his methods and the reasons he gave in his report.

Holford concludes, based on his biostatistical analysis of the data, that McGovern contains a serious tabulation error that renders the 3.8 odds ratio, and the claimed “strong” association it purportedly demonstrates, overstated and erroneous. As Holford explains, the data show one fewer infection in the Bair Hugger period and one additional infection in the HotDog period than reported in the published McGovern study. Comparing the spreadsheets produced by Augustine and Dr. McGovern explains the origin of the discrepancy. All the procedures identified as taking place during the Bair Hugger period are identified as “FAW” in the data produced by McGovern, but only three of the four procedures occurring during the HotDog period are coded as “CWB.” One procedure (September 15, 2010) is coded as “FAW,” meaning that it was counted as a Bair Hugger infection. And that is the problem. September 15, 2010 was long after Northumbria fully transitioned from Bair Hugger to HotDog and was the exact midpoint of what McGovern reported as the HotDog period. At his deposition, Dr. McGovern

could not explain why a procedure occurring in the middle of the HotDog period was coded as “FAW” (*i.e.*, as a Bair Hugger infection). McGovern Dep. II at 481:6–484:6.

Holford’s analysis explains how this occurred. All the initial entries’ dates on the data produced by McGovern are presented in the standard British format of date, month, year. These are identical to the same entries on the data produced by Augustine for the corresponding procedure. On the final pages of the McGovern-produced data, however, where the “FAW” or “CWB” coding is entered, the procedure dates are repeated but entered in the standard American format of month, date, year. Per Holford, this suggests that the warming-device coding likely occurred on the U.S. side of the Atlantic, after Reed provided the original dataset to Augustine employee Mark Albrecht for statistical analysis. Albrecht apparently coded the warming devices on an Excel spreadsheet that had culled out only those procedures that resulted in infections. Whether by error or by intention, the September 15, 2010 procedure was miscoded as “FAW” (Bair Hugger) when it should have been “CWB” (HotDog).

3. McGovern’s co-author Mike Reed recalled the tabulation error slightly differently, and Holford also accounted for the possibility that Reed’s recollection was accurate.

Contrary to Plaintiffs’ arguments, Holford is not alone in finding tabulation errors in the McGovern data. McGovern’s authors knew of tabulation errors and that the reported odds ratio of 3.8 was substantially higher than the actual data supported. Reed testified that there was an additional infection in both the Bair Hugger and HotDog groups that had, despite his communications with Albrecht, not been corrected prior to publication. Reed Dep. 42:23–44:9. Plaintiffs dismiss this testimony as “unexplained,”

as though the sworn testimony of McGovern's senior author must be ignored simply because he did not offer an explanation that satisfied Plaintiffs.⁵

The data produced by McGovern and Augustine suggest that Reed's recollection of the nature of the tabulation error is not quite accurate: there was actually one fewer Bair Hugger infection, rather than one additional infection in both the Bair Hugger and HotDog groups. For the sake of completeness, Holford accounted for the possibility that Reed's recollection was correct. Holford performed a biostatistical analysis assuming the accuracy of Reed's recollection, in addition to analyzing the data assuming the September 15, 2010 miscoding. Based on Reed's testimony, he calculated an odds ratio of 2.86 with a wide confidence interval just barely above statistical significance.⁶ DX1, Holford Rpt. 3 n.1. Thus, Reed's testimony that there should have been one additional infection counted in both periods also leads to a result markedly different from McGovern's published findings.

Plaintiffs also ignore the written communications between Reed and Albrecht, which show that they were aware that data in the published article differed from an updated dataset that Reed sent to Albrecht. DX17, 5/27/2012 Emails between Albrecht and Reed. In these private communications that were never disclosed to the peer

⁵ Plaintiffs incorrectly assert that Holford agreed with Reed that "adding or subtracting a single DJI does not impact the 'scientific, human, or economic' significance of the published study." Pl. Holford Mem. 11. In this context, Holford was testifying about the general concept of statistical significance, not the impact of the tabulation errors. DX2, Holford Dep. 84:8-11. ("Q. *Statistical significance* is not equivalent to scientific, human, or economic significance; correct? A. Correct.") (Emphasis added).

⁶ Assuming the miscoding error, Holford calculated an odds ratio of 2.76, which was not statistically significant. Holford Rpt. 1.

reviewers, Albrecht recalculated the odds ratio with the updated data and found that it was 2.98, not 3.8 as reported. *Id.* Whether McGovern should have reported the odds ratio as 2.98 (as Albrecht conceded in his e-mail exchange with Reed), 2.86 (based on Reed's testimony), or 2.76 (based on the apparent miscoding of the Sept. 15, 2010 procedure) is beside the point. The bottom line is that the odds ratio reported in the published study and relied upon by Plaintiffs' experts is, by its authors' own admissions, **wrong**. Tabulation errors occurred, the tabulation errors resulted in an inflated odds ratio, and neither McGovern's authors nor Plaintiffs' experts offer any correction or explanation for this major flaw.

Holford clearly articulated the assumptions he made and the reasons he made them in analyzing the implications of Reed's testimony. Holford Rpt. 2–6. Plaintiffs' own expert Dr. Samet acknowledged that Holford performed this analysis correctly:

Q. Do you have any reason to think Professor Holford screwed up the calculations that he did there [in footnote 1, which addresses the implications of Reed's testimony]?

A. Oh, he certainly did the calculations correctly.

Samet Dep. 126:3-7.

In sum, Holford reliably applied biostatistical analysis (the same subject that he teaches) to find that the data underlying the McGovern study were incorrectly analyzed by the authors and that the published results are inaccurate. It is this type of serious error uncovered during discovery that caused the court in the Viagra MDL to revisit its initial *Daubert* decision (the decision Plaintiffs repeatedly cite without noting its subsequent history). The *Viagra* court initially denied the defendants' motion to exclude Plaintiffs'

general causation experts because they relied upon a peer-reviewed, published study. But following additional discovery that exposed data discrepancies, it revisited that decision and granted the defendants' renewed motion. *See Viagra II*, 658 F. Supp. 2d at 950. Summary judgment for the defendants followed.

The same result should follow here. Peer reviewers who had no knowledge of the tabulation errors or access to the underlying raw data could not have detected the errors identified by Holford. Plaintiffs' oft-repeated incantation of "peer review" as a shield for these serious errors cannot serve as a basis for excluding Holford's analysis or Borak's opinions based on that analysis. As the *Viagra* court noted, "[p]eer review and publication mean little if a study is not based on accurate underlying data." *Id.* at 945.

4. Holford and Borak demonstrate the fatal impact of disclosed confounders on McGovern's reliability.

The second fatal flaw with McGovern, as identified by both Holford and Borak, is the failure to account for the impact of confounders. McGovern's authors disclose only two potential confounders. McGovern 1543. When even the two disclosed confounders are accounted for, the "association" between the Bair Hugger system and increased infection risk vanishes.⁷ Plaintiffs' selective and inaccurate quotations from the scientific literature fail to undermine Holford and Borak's methods or conclusions.

⁷ Mark Albrecht, the Augustine employee who performed the data analysis for the McGovern study, conceded this point: there is no significant difference in infection rates when one controls for the disclosed confounders. DX13, Albrecht Dep. 200:9–205:18.

i. Confounder 1: Rivaroxaban

As described in McGovern, at the beginning of the Bair Hugger period, all patients were given tinzaparin as thromboprophylaxis (to prevent blood clots). Several months after changing the antibiotics, but prior to switching to HotDog, the hospital switched the thromboprophylaxis regimen to rivaroxaban. However, after seven months, the complication rate from rivaroxaban was so high that the hospital switched back to using tinzaparin. Reed Dep. 92:8–95:15. The full seven months of rivaroxaban use occurred during the Bair Hugger period; indeed, it coincided with the end of the Bair Hugger period. Thus, all HotDog patients had the benefit of tinzaparin.

Rivaroxaban's impact on the infection rate during the Bair Hugger period is critical. Holford demonstrates that a major spike in infections occurred during the seven months rivaroxaban was used. In response, Plaintiffs argue that rivaroxaban made no difference on the infection rates. They base their argument primarily on another paper published by Reed and other colleagues on their seven-month experience with rivaroxaban. *See* DX18, Jensen et al., "Return to Theatre Following Total Hip and Knee Replacement, Before and After the Introduction of Rivaroxaban," 93-B(1) *J. Bone & Joint Surgery* 91 (2011).

Plaintiffs have seriously misrepresented Jensen's findings. Based on Plaintiffs' description of Jensen, one might conclude the study found no difference in infection rates between the six-month tinzaparin period and the seven-month rivaroxaban period (Bair Hugger was used in both periods). But that's not what Jensen says. As reported in Jensen, the overall deep infection rate for the tinzaparin period was 1 percent versus 2.5

percent for the rivaroxaban period. Jensen at 93. These numbers were based on five infections in the tinzaparin period versus 14 infections in the rivaroxaban period.⁸ *Id.* Given the small numbers, the difference between 1 percent and 2.5 percent fell short of statistical significance ($p = 0.102$). *Id.* As Reed testified, “It wasn’t far off significance, I will give you that, but if you -- we couldn’t link rivaroxaban to infection.” Reed Dep. 106:21-23.

Plaintiffs nonetheless assert that Jensen finds that the infection rates are “similar.” Pl. Holford Mem. 24. That is not what Jensen says either. The “similarity” that Jensen discusses on page 91 is not between the deep joint infection *rates*; rather, it is the “*proportion* of patients who return to theatre and became infected” that remained “similar.” Jensen at 91 (emphasis added). Jensen explains that 55.5 percent of the tinzaparin patients who returned to theatre had a deep infection, compared with 66.63 percent of the patients who returned to theatre in the rivaroxaban group. *Id.* It was these percentages that Jensen characterized as “similar,” not the overall rate of infection for each group.

Nor did Jensen say that the deep joint infection rates in the tinzaparin and rivaroxaban groups were “similar” on page 93. There, Jensen notes: “Our rate of infection increased from 1% to 2.5% following the introduction of rivaroxaban. An infection rate of 1% is similar to that reported in the literature following hip and knee

⁸ As Defendants’ expert Dr. Mont noted, regardless of “statistical significance,” no hospital would disregard a jump from five infections in a six-month period to 14 infections in a seven-month period. DX19, Mont Rpt. 15–16. Northumbria did not disregard it, and reverted to tinzaparin after only seven months.

replacement.” *Id.* at 93. In other words, Jensen noted that the 1 percent rate in the tinzaparin group was “similar” to the infection rate reported in the general literature.

Holford carefully analyzed Jensen. He noted that Jensen had a shorter window for infection surveillance than McGovern (30 versus 60 days) and included not only elective joint surgery, but trauma surgery as well. Holford Rpt. 5. McGovern excluded trauma patients because of the much higher rate of infection. Holford Rpt. 5; *see* Reed Dep. 60:6–63:15. Knowing the dates of tinzaparin use versus rivaroxaban use as reported in both Jensen and McGovern, Holford demonstrates that four of the infections reported in the Bair Hugger cohort during the period when rivaroxaban was used occurred between 30 and 60 days following surgery. Thus, those four infections were not included in the 14 infections reported in Jensen’s rivaroxaban group. Holford applies McGovern’s surveillance period (60 days) and its trauma-exclusion criterion to Jensen’s dataset, which results in a highly significant impact of rivaroxaban on deep joint infections. Holford Rpt. 5.

Plaintiffs accuse Holford of “mixing and matching” data here. Pl. Holford Mem. 1. But it is Plaintiffs who “mix and match.” Plaintiffs attempt to compare the apple of Jensen’s 30-day surveillance period to the orange of McGovern’s 60-day period, misrepresent what Jensen said, disregard the outcome-determinative difference that the surveillance period had, and declare that rivaroxaban could not possibly be a confounder.⁹ Holford’s reliable statistical analysis exposes these misrepresentations.

⁹ Tellingly, in response to a peer reviewer’s comment about the 60-day surveillance period, McGovern’s authors noted that “if we had chosen a 30 day cutoff, we would have

Plaintiffs also cite four studies comparing the efficacy of rivaroxaban to other anticlotting drugs (but, significantly, *not* tinzaparin). Pl. Holford Mem. 24–25; Pl. Borak Mem. 15. Per Plaintiffs, these four studies, known as the “RECORD studies,” demonstrate that rivaroxaban does not increase deep joint infections over other anti-clotting agents. *Id.* But that issue simply was not addressed in the RECORD studies. As Borak testified, the RECORD studies did not compare rivaroxaban to tinzaparin, nor did they define joint infections as an endpoint. DX4, Borak Dep. 206:24–208:14.¹⁰ Jensen made the same point:

The RECORD trials were deficient in their lack of measurement of surgical outcomes such as wound healing, drainage, *infection*, range of movement, and chronic pain. This led one of the authors of the RECORD paper to later state that he would not recommend it (rivaroxaban) for his patients.

Prolonged wound drainage after lower limb arthroplasty is associated with infection, longer hospital stay, re-operation, and a subsequent increase in the economic burden on the national resources. We could not find any reports focusing on the potential wound complications associated with the use of oral factor Xa inhibitors such as rivaroxaban.

Jensen at 93 (emphasis added). Jensen’s authors were not the only ones to note that the RECORD studies failed to measure infection. As a recent study noted:

[The RECORD trials] demonstrated rivaroxaban to have superior efficacy compared to enoxaparin . . . in preventing

missed 10.9% of deep joint infections.” Response to Reviewer Comments at 2557. That 10.9% equates to four additional infections, all of which occurred in the rivaroxaban period.

¹⁰ Plaintiffs say that Borak did not consider the RECORD studies, but he plainly did. *See also* DX20, Wenzel Dep. 259:23–260:3.

VTE [venous thrombo-embolism] with no significant increase in the major bleeding risk. Although these studies favored rivaroxaban in terms of reduced VTE rates, wound infection and subsequent re-operation were outcomes that were not fully evaluated by the record trial design.

DX21, Brimmo et al., “Rivaroxaban Use for Thrombosis Prophylaxis is Associated with Early Periprosthetic Joint Infection,” 31(6) *J. Arthroplasty* 1295 (2016). Brimmo found the deep joint infection rate to be 0.2 percent in the control group versus 2.5 percent in the rivaroxaban group. *Id.* at 1295. The study concluded that the “use of rivaroxaban for thromboprophylaxis led to a significantly increased incidence of deep SSI in a continuous series of patients undergoing primary THA and TKA in a single institution.” *Id.*

Brimmo found that “[p]atients who received rivaroxaban had a relative risk of infection of 10.7 (95% CI: 1.2-94.9; $P = .021$).” *Id.* 1297. This highly significant finding that rivaroxaban increases the risk of deep joint infection by more than ten-fold is, of course, further evidence that the use of rivaroxaban for seven months of the Bair Hugger period (and the dramatic increase in infections in that period) is a confounding factor.

In sum, the scientific literature supports Holford and Borak and provides no support to Plaintiffs’ attack on the admissibility of their opinions.

ii. Confounder 2: Antibiotic Prophylaxis

The other possible confounder disclosed by McGovern, and addressed by Holford and Borak, is antibiotic prophylaxis. As McGovern notes, patients at the beginning of the Bair Hugger period were given only one prophylactic antibiotic: gentamicin. Several months into the Bair Hugger period, a second antibiotic, teicoplanin, was added to the

regimen. This regimen remained constant for the remainder of the Bair Hugger period and throughout the HotDog period.

Plaintiffs argue, against Holford and Borak, that the change in the antibiotic prophylactic protocol during the Bair Hugger period could not be a confounding factor because two published studies, Melling and Hickson, “find no meaningful difference in the effect of different antibiotic regimens and DJI rates.” Pl. Borak Mem. 11, citing DX22, Melling et al., “Effects of Preoperative Warming on the Incidence of Wound Infection after Clean Surgery,” 358 *Lancet* 876 (2001); DX23, Hickson et al., “Prophylactic Antibiotics in Elective Hip and Knee Arthroplasty,” 4(11) *Bone & Joint Res.* 181 (2013). Again, the literature cited by Plaintiffs does not in any way demonstrate that Holford and Borak applied an improper methodology.

Melling, one of the two “gold standard” randomized clinical trials examining the impact of warming on surgical site infections, discussed the issue of antibiotics in a background section only. It was not a study about antibiotics, nor was it a study about deep joint infections in arthroplasties. Melling 876. Melling does state that “the benefits of prophylactic antibiotics in reducing infection rates after clean surgery remain unclear.” *Id.* But Plaintiffs ignore the robust evidence that has evolved in the 16 years since Melling was published that has found, with respect to deep joint infections, a clear benefit from the use of properly-timed pre-incision prophylactic antibiotics. This literature strongly supports Holford and Borak’s conclusions and underscores the reliability of their approach.

Hickson makes this point clearly. After describing prophylactic antibiotics as a “key intervention” to reduce rates of deep joint infection, Hickson cites a 2008 pooled analysis of seven studies that found the administration of prophylactic antibiotics reduced the relative risk of wound infection by 81 percent. Hickson 182.

Plaintiffs also quote Hickson’s statement that “there is no clear benefit to using one particular agent/regimen.” Pl. Borak Mem. 11. They say this means that combination of teicoplanin and gentamicin (the drug regimen to which Northumbria switched during the Bair Hugger period and maintains to this day) is no more effective than gentamicin alone, the antibiotic protocol in place during the first seven months of the Bair Hugger period. *Id.* 11–12. This is an egregious misreading, and contradicted by language on the very same page. Hickson notes that while there is a “large body of evidence” supporting prophylactic antibiotics, “[t]here is no evidence for the use of systemic gentamicin as prophylaxis in primary elective THA and TKA surgery.” Hickson 186.

Given that Reed co-authored the Hickson study, this unambiguous conclusion was almost certainly informed by Reed’s own experiences at Northumbria during the period of the McGovern study. Reed’s experiences are also detailed in another published paper Plaintiffs have not cited: DX24, Sprowson et al., “Changing Antibiotic Prophylaxis for Primary Joint Arthroplasty Affects Postoperative Complication Rates and Bacterial Spectrum,” 11 *Surgeon* 20 (2013). Sprowson was the senior author on Hickson while Reed was the senior author on Sprowson.

Sprowson examined complications, including wound infection, in joint arthroplasty at Northumbria from January 2002 to February 2009, a period that involved a switch from cefuroxime to gentamicin. Sprowson at 20. Cefuroxime had been used at Northumbria with great success; however, reducing clostridium difficile-associated diarrhea (CDAD) became a national priority. *Id.* at 20, 21. As part of Northumbria's CDAD improvement plan, antibiotic prophylaxis for primary arthroplasty was changed from cefuroxime to gentamicin in October 2007. *Id.* at 21. As we know from McGovern, Northumbria changed from gentamicin-only to gentamicin-plus-teicoplanin in March 2009.

Sprowson provides remarkable insight into the issues at Northumbria prior to the switch to HotDog, and sheds light not only on the importance of the right combination of antibiotics in preventing deep joint infections, but also the complete lack of any involvement of the Bair Hugger in causing them. In the nearly six years of cefuroxime use examined by Sprowson's authors, involving 6,094 patients undergoing primary knee and hip arthroplasty, the return to theatre (RTT) for proven infections was only 0.66 percent. *Id.* In the 17 months of gentamicin-only use reviewed in the Sprowson paper, involving 2,101 patients receiving a single dose of gentamicin, the RTT rate for proven infection jumped to 1.52 percent. *Id.* This difference was statistically significant ($P < 0.01$). *Id.* Reed and his co-authors concluded:

[G]entamicin 4.5 milligrams per kilogram alone should not be used as prophylaxis for primary joint arthroplasty as it does not reduce CDAD significantly but increases the risk of postoperative complications. We have changed our

prophylaxis to low dose gentamicin (3 mg. per kg combined with teicoplanin 400 mg. given once).

Id. at 20.

Sprowson is revealing for several reasons. First, it examined infection rates at the Trust that includes the hospital that McGovern studied. Second, a portion of the patients Sprowson examined included the same patients who comprised the first seven months of the Bair Hugger period in the McGovern study. All of the more than 8,000 surgeries in Sprowson, however, occurred when the Trust was using Bair Hugger exclusively. When using cefuroxime, the infection rate was only 0.66 percent. After switching to gentamicin, the infection rate nearly tripled, to 1.52 percent. This statistically significant increase prompted a change in the antibiotic regimen seven months into the Bair Hugger period.

Sprowson refutes Plaintiffs' attack on Holford and Borak's conclusion that the antibiotic regimen is an important confounder. Using gentamicin alone nearly *tripled* the infection rate over another antibiotic, prompting Northumbria to change its antibiotic regimen. Further, when the more effective antibiotic cefuroxime was used in conjunction with the Bair Hugger, the joint infection rate was a mere 0.66 percent, markedly lower than the rate "achieved" by switching from Bair Hugger to HotDog.¹¹

¹¹ Tinzaparin had been the thromboprophylaxis at Northumbria from October 2006 until a brief and disastrous switch to rivaroxaban, prompting a return to tinzaparin. DX6, Khan et al., "Reduced short-term complications and mortality following Enhanced Recovery primary hip and knee arthroplasty: results from 6,000 consecutive procedures," 85(1) *Acta Orthopaedica* 26, 27 (2014).

Sprowson's importance does not end there. Plaintiffs have argued, and Reed has contended, that July 2008 was selected as the start date for McGovern because the surveillance prior to July 2008 was not sufficiently robust as to be reliable. Sprowson demonstrates, however, that Reed had no difficulty accessing necessary infection data when he wanted:

Data was collected on all patients who returned to theatre (RTT) after their primary arthroplasty up to 1 year. The medical records of these RTT patients were scrutinized to determine whether the procedure was related to their joint replacement. The microbiological records of those patients were also examined to determine if samples had been taken at the time of surgery. If the microbiological and clinical findings fulfilled the HPA [Health Protection Agency] definition criteria, this was recorded as RTT. The infecting organism and sensitivity profile were recorded.

Sprowson at 21. Thus, contrary to Plaintiffs' assertion, Reed could access joint-infection data for primary joint arthroplasty patients prior to July 2008. These data were sufficiently robust for Reed to publish the Sprowson article in the *Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland*. Sprowson was submitted for publication less than three months after McGovern was published.

5. Holford's analysis demonstrates that the chosen start date affects McGovern's results and thus its reliability.

What is the significance of the fact that Reed had robust infection data for procedures prior to July 2008? The answer is found in the start date of the underlying data. These data begin in October 2007, the month when Northumbria switched from cefuroxime to gentamicin.

These earlier data are the basis for one of the most important portions of Professor Holford's biostatistical analysis. If McGovern had looked at the Bair Hugger period starting in October 1, 2007, rather than July 1, 2008, the difference in infection rates between Bair Hugger and HotDog is, according to Holford, "not close to being significant." Holford Rpt. 5.¹² Further, as of July 1, 2008, the start date ultimately chosen for the Bair Hugger period, the infection rate was already beginning to rise. *Id.*

Holford conducted a statistical analysis of the data using start dates of each successive month, beginning in October 2007. As demonstrated on Holford's Figure 3, using the same chi-square method Albrecht used, significance is not achieved for any month prior to July 1, 2008. *Id.* at 3. And if the start date had been two months *later*, statistical significance would not be achieved either.¹³

Holford's statistical analysis cannot get inside the minds of the McGovern co-authors, who performed the data analysis for the McGovern study. But it is certainly relevant in evaluating McGovern's reliability that its co-authors (who included Augustine employee Mark Albrecht) selected as the start date for the Bair Hugger period the date that would get them to five percent significance *after* they had the data.

The early drafts of the McGovern study suggest that the start date used in the published study was selected to achieve statistical significance. The first draft states,

¹² Specifically, $P = .2179$ using Fisher's exact test, with an odds ratio of 2.12 (95% confidence interval of 0.75-6.00). Holford Rpt. 5.

¹³ Other than July, August, and October 2008, the P value is below five percent significance level for any start month until February 2009, at which point it remains above the five percent significance level.

“Joint sepsis data was collected for all orthopedic operations performed in the hospital during the 2-year period prior to the study, with dates comprising *9/1/2008 to 9/1/2010*.” DX25, McGovern Drafts at 2206 (emphasis added). In the tenth draft, however, the authors expanded the Bair Hugger period by moving the start date from September 2008 back to July 2008, the date used in the published paper that achieved (just barely) statistical significance. *Id.* at 2400–2422. As Holford demonstrates, it took moving the start date to achieve even minimal significance. It is not hard to hit a statistical bullseye when you shoot an arrow first, and then paint the target around the arrow.

How the authors displayed the infection data graphically is also telling. In the early version of Figure 7, the infection data are plotted on a graph showing percentage of infections on the X axis, a timeline on the Y axis, and a “jitter plot” reflecting individual infections across the top.

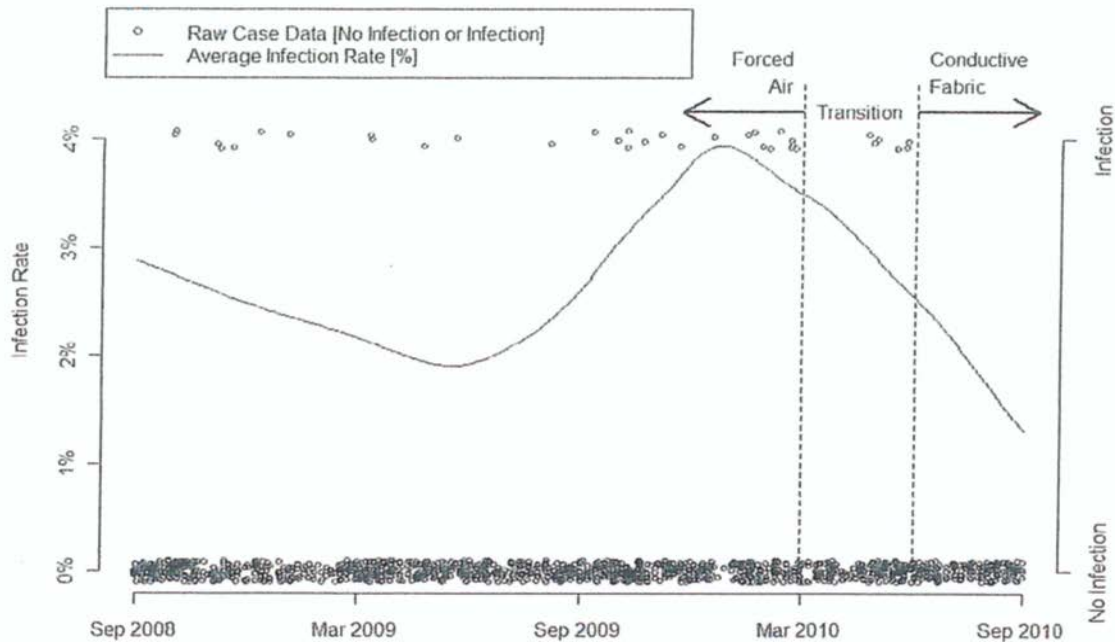


Figure 7: Infection data for n=1290 joint replacement cases with the outcome plotted on the right hand axis (data is jittered to avoid overprinting). A moving average of infection rate was plotted on the left hand axis. The change from forced air to conductive fabric patient warming in the orthopedic theaters is identified along with the transition period where both systems were used.

McGovern Drafts at 2218. The start date in the first version, consistent with the text of the first draft, was September 2008; in the published version, it was July 2008. The infection data are also plotted in a manner that shows the rising and falling rates over time. This original version of Figure 7 shows an infection rate at the beginning of the period of just under three percent, declining over a period of several months to a low of just under two percent, then rising sharply in the last few months to four percent before drifting back down following the switch from Bair Hugger to HotDog. Thus, the graphic depiction of the infection data during the Bair Hugger period in this original version of Figure 7 would have alerted the reader to the fact that the infection rate varied widely during this period, and—critically—that there was a major spike in infection rates in the

brief few months immediately preceding the transition to the HotDog warming unit. These same months coincided with the hospital's use of rivaroxaban.¹⁴

By the third draft, the graphic depiction in Figure 7 had changed to eliminate the rolling average. McGovern Drafts at 2262. Significantly, in the tenth draft, infection data were added for the HotDog period. The text reflected an infection rate of 1.08 percent for the HotDog period. McGovern Drafts at 2410.¹⁵ The jitter plot for Figure 7 reflected four infections during the HotDog period. In addition, the start date moved from September 2008 back to July 2008.¹⁶

¹⁴ No mention was made of rivaroxaban or any other potential confounders in the early drafts. References to rivaroxaban and the antibiotic regimen were added later.

¹⁵ No odds ratio was presented; however, given the two infection rates (1.08 versus 3.10), it is simple arithmetic: 2.87, almost exactly the number that Holford calculated. The P value is given as 0.043 (McGovern Drafts at 2409), a P value that, while meeting the general convention of what is considered to be statistically significant, does so just barely. (An odds ratio of 2.87, with a minimally significant P value of 0.048, could hardly be described as evidence of a “strong association.”)

¹⁶ The tenth draft is further evidence that the data produced by Augustine are the “real” McGovern data. The total number of HotDog procedures identified in the tenth draft is 372, one *more* than in the published paper (371). McGovern Drafts at 2410. The infection rate for this group was shown to be 1.08 percent. 1.08 percent of 372 is 4—precisely the number of infections reflected in the data produced by Augustine and analyzed by Holford. Likewise, the number of Bair Hugger procedures in the tenth draft is 1060, one *fewer* than the number in the published version, and is consistent with Holford's analysis of the tabulation error.

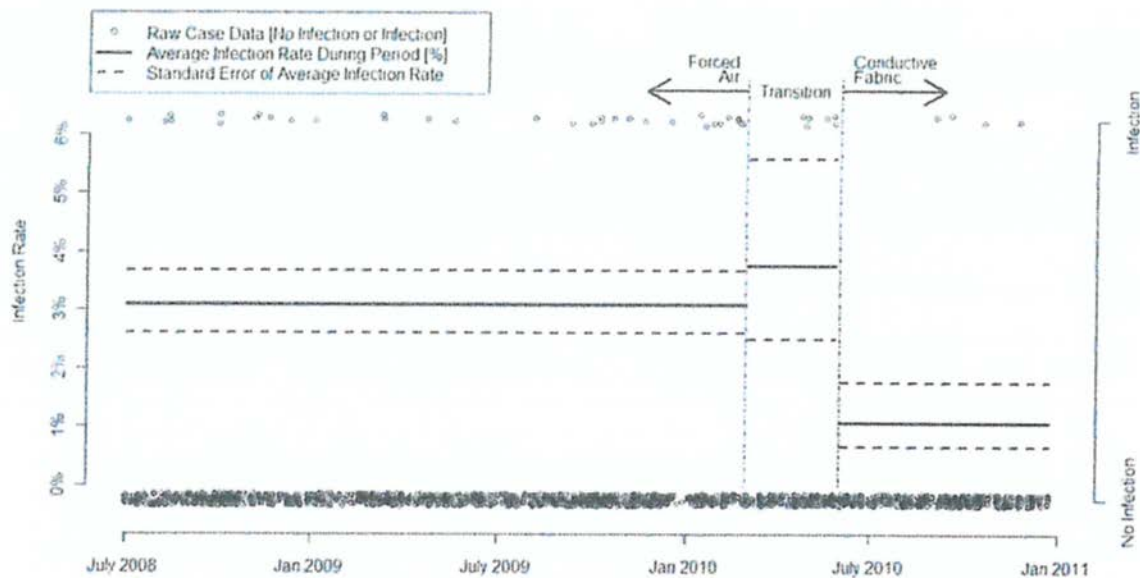


Figure 7: Infection data for n=1597 joint replacement cases with the outcome plotted on the right hand axis (data is jittered to avoid overprinting). Average infection rates for each period (Forced air, Transition, or Conductive Fabric) are plotted on the left hand axis. Standard error for the average infection rate was estimated using logistic regression.

McGovern Drafts at 2419. Ultimately, the published version of Figure 7 reflected the flat line averages introduced in the third draft, concealing the rivaroxaban spike, and reduced the reported infections in the HotDog period from four to three, reducing the infection rate, as shown below:

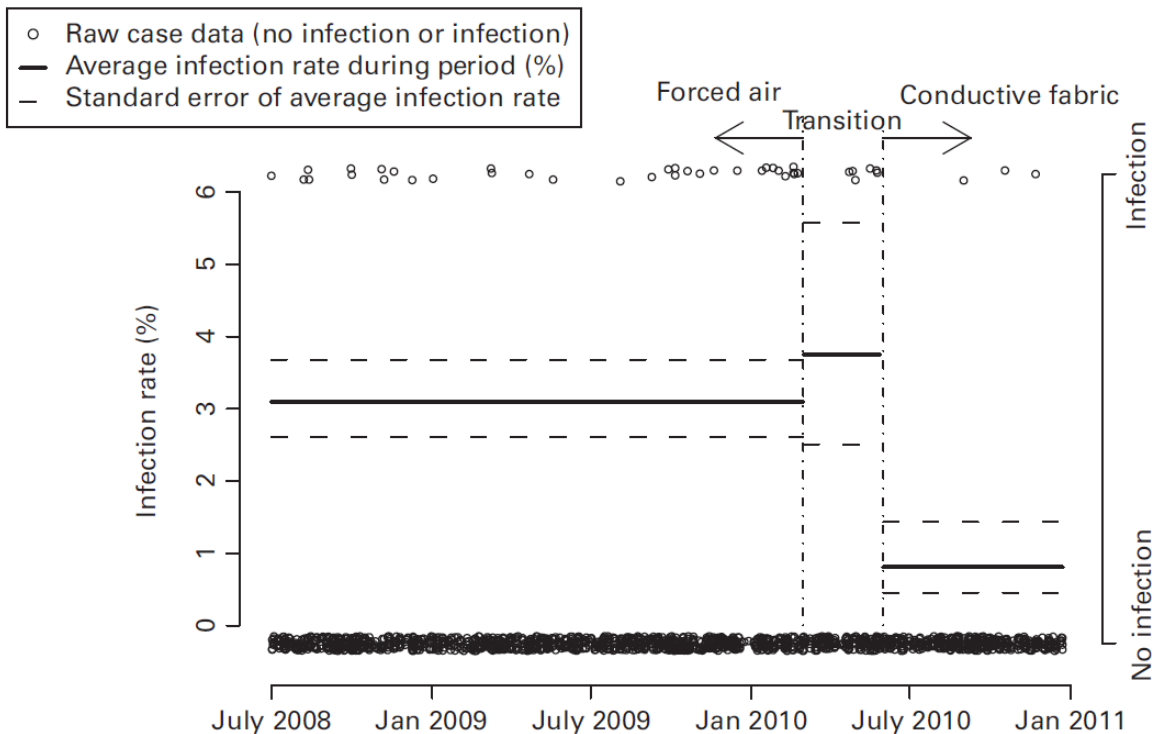


Fig. 7

Graph showing time-based trends of joint sepsis rates for hip and knee replacement cases. The outcome of each individual case is plotted on the right-hand axis (data are jittered to avoid overprinting). The infection rates for each period (forced-air, transition or conductive fabric) are plotted on the left-hand axis. Standard error of the mean was estimated using logistic regression.

McGovern at 1543.

By adjusting the start date and reclassifying one of the four HotDog infections as a Bair Hugger infection, Albrecht and his co-authors could strengthen the apparent association and its statistical significance and make it less likely that the medical community would attribute this apparent association to confounders. The change would, of course, also more strongly advance the “agenda” of Albrecht’s employer, Augustine.¹⁷

¹⁷ Augustine, Albrecht, and their co-authors took great pains to “hide [the] agenda” of their publications and to appear “impartial.” See DX26, 2/16–22/2011 Emails from McGovern to Albrecht at 5 (“You did a good job of hiding the ‘agenda’ and making this look much more impartial”); DX27, BHS Presentation Outline at 1 (“This makes it look impartial and hides our agenda, so to speak...”); DX28, 1/23/2010 Email from Albrecht to

It is not necessary for Defendants to demonstrate, or the Court to find, that mistabulating infection data, selecting an earlier start date, or changing the graphical representation of the data were intentional or manipulative. These facts do, however, corroborate Holford and Borak's conclusion that McGovern is fundamentally flawed and should not have been relied upon by Plaintiffs' experts. As the 7th Circuit has noted:

[A] high significance level may be a misleading artifact of the study's design; and there is always the rush that the party's statistical witness ran 20 regressions, one and only one of which supported the position and that was its only one presented, though, in the circumstances, it was a chance result with no actual evidentiary significance (Careful pretrial discovery by the other party should unmask this trick.)

Kadas v. MCI Systemhouse Corp., 255 F.3d 359, 362 (7th Cir. 2001).¹⁸

Holford, Borak, and the discovery taken in this litigation have unmasked McGovern's trick. If the Court does not exclude McGovern and Plaintiffs' experts, Holford and Borak's relevant, reliable, and objective analysis will be—at a minimum—invaluable to the trier of fact in understanding how, through data manipulation, the McGovern authors were able to generate the “strong association” that now underpins Plaintiffs' case.

Leaper at 1 (“we need to be critically careful that this document appears to be impartial”); *see also* Albrecht Dep. 310:19–324:13.

¹⁸ The response to a reviewer's comment about the surveillance period is further evidence that the authors selected key parameters *after* analyzing the data. *See supra* n. 9.

II. **BORAK APPLIED A VALID METHODOLOGY TO IDENTIFY ADDITIONAL UNDISCLOSED CONFOUNDERS IN MCGOVERN.**

Borak identifies additional undisclosed confounders as further reasons why McGovern is not a reliable foundation for Plaintiffs’ medical experts’ opinions, and appropriately criticizes Plaintiffs’ experts for failing to account for them. In particular, he notes that a major infection control initiative was implemented by Northumbria during the time of the study that primarily benefited HotDog patients.

A. **Massive Infection Reduction Measures Were Undertaken at the Hospital Studied by McGovern—Measures that Disproportionately Benefited the Non-Bair Hugger Patients.**

As Borak notes based on his review of McGovern co-author Reed’s testimony, massive efforts were being undertaken by the Northumbria Healthcare NHS Trust at the time of the McGovern study to decrease its infection rates. Reed Dep. 78:3–79:2, 108:25–109:11, 112:18–113:22, 114:7-25, 115:11–119:6. These efforts are also described in a published article authored by the senior nurse and leader of the surgical site infection surveillance team at Wansbeck. *See* DX29, J. Gillson & G. Lowdon, “Implementing effective SSI Surveillance,” *Clinical Scis. J.* 71 (Oct. 2014). Virtually all these changes were implemented during or at the end of the Bair Hugger period (July 2008 to February 2010), meaning that patients in the HotDog period benefitted from infection control protocols that had not been in place for the majority of patients in the Bair Hugger period. *See id.*, Figure 2; DX3, Borak Rpt. 10–12.

Reed and his hospital colleagues disclosed these changes publicly—just not in the McGovern paper itself. In addition to the Gillson paper, Reed and his colleagues

published various studies on changes impacting infection control procedures at Northumbria: for example, in the Jensen and Sprowson papers, and a peer-reviewed article (“Refaie”) specifically discussed best practices for reducing joint infections based on Northumbria’s experience. *See* DX30, R. Refaie, “Prevention of Periprosthetic Joint Infection,” 3(3) *J. Trauma & Orthopedics* 50 (Sept. 2015).

Refaie discusses prewarming, a practice instituted at Northumbria at the same time the HotDog was introduced—so no Bair Hugger patients would have had the benefits of prewarming, while all HotDog patients would have:

Prewarming of patients before theatre is a proven strategy for preventing hypothermia intraoperatively and in recovery. A large RCT from the UK published in the Lancet showed that prewarming reduced the risk of infection by around 65% in clean surgery. Despite this, prewarming is still not widely adopted in UK centers.

Refaie at 51. Refaie also describes Northumbria’s experience in implementing screening for Methicillin Susceptible *Staph. aureus* (MSSA)—a practice instituted during the Bair Hugger period:

After MSSA screening and decolonization was introduced in one NHS joint replacement unit, MSSA infections reduced from 0.84% to 0.26% – the caveat being there were other infection prevention methods implemented during this time period.

Id. This statement’s footnote demonstrates that Northumbria was the NHS joint replacement unit to which Refaie was referring. Refaie thereby confirms that MSSA infection rate dropped by more than two-thirds when MSSA screening and decolonization was implemented at Northumbria. Because MSSA screening was not

implemented until January 2010, most Bair Hugger patients in the McGovern study would not have had the benefit of MSSA screening. Most Bair Hugger patients had their surgery conducted at a time when the MSSA infection rates were themselves nearly one percent. By contrast, every one of the HotDog patients would have benefited from MSSA screening and decolonization which, according to Rafaie, caused MSSA infection rates to plummet. Significantly, *none* of the infections in the HotDog period were *staph. aureus*, while nearly a third of the Bair Hugger infections pre-MSSA screening were *staph. aureus*.

B. Plaintiffs Fail to Rebut Borak’s Conclusion that These Infection Control Measures Undermine McGovern’s Validity.

Borak discusses several of these undisclosed infection control initiatives and the potential impact that they have on the reliability of the McGovern study. Borak Rpt. 10–12. He explains that these confounders plague the McGovern study, dramatically undermining the validity of its conclusions, and Dr. Samet’s reliance upon it. *Id.* 19. Plaintiffs’ rejoinder to Borak boils down to this: unless an infection control practice has been conclusively demonstrated to have a statistically significant impact on *overall* infection rates by itself, it cannot be a real confounder.

MSSA infection screening demonstrates the inapplicability of Plaintiffs’ position. Screening for MSSA, and decolonizing those patients who test positive for MSSA prior to their joint replacement surgery, would not have any impact on infections caused by bacteria *other than MSSA*. The impact of a single infection control strategy directed at one specific type of bacteria may well not show up as “statistically significant” in terms

of its infection reduction for *all* bacterial infections, but its impact may well, as in the case of *staph. aureus* at Northumbria, be highly significant for that particular bacterial infection.

Both Borak and Holford discuss the reasons why a confounder need not be demonstrated to have a “statistically significant” impact on an outcome for it to nonetheless be considered a confounder. Holford Dep. 280–81; Borak Dep. 148. This is a basic premise accepted in biostatistics and epidemiology and, in fact, conceded by Plaintiffs’ own epidemiology expert, Dr. Samet. Samet Dep. 50:8-9. Put simply, Plaintiffs’ contrary position finds no support in the scientific literature or their own experts’ opinions.

C. The Remainder of Plaintiffs’ Attack on Borak Consists of Misrepresentations of His Report and His Methods.

The remainder of Plaintiffs’ attack on Borak is based on mischaracterizations of his report and testimony. The Addendum addresses them point for point. In sum, Borak’s opinion is the product of a careful and meticulous scientist applying widely accepted methodologies (the same methodologies he teaches to his public health and medical students at Yale), and it would help the jury understand McGovern and the lack of scientifically convincing support for Plaintiffs’ experts’ opinions.

III. PLAINTIFFS’ ARGUMENT THAT MCGOVERN IS RELIABLE *PER SE* IS CONTRARY TO CASE LAW.

Holford’s and Borak’s analyses and conclusions are highly relevant because they reinforce what Defendants had already established through discovery and depositions of McGovern’s co-authors: that its flaws vitiate its reliability. Plaintiffs argue, however,

that even if Holford and Borak are correct, their opinions are irrelevant because the McGovern study is reliable as a matter of law. Because (Plaintiffs assert) McGovern is “pre-litigation research” and “peer reviewed,” it is admissible *per se*. As discussed below, Plaintiffs’ *per se* rule finds no support in case law and is, in fact, contrary to the cases they cite.

A. Plaintiffs’ Cite to the *PPA* Case Only Underscores Why McGovern Is Not Reliable and the Importance of Borak and Holford’s Testimony.

Plaintiffs rely heavily on *In re Phenylpropanolamine (PPA) Prods. Liab. Litig.*, 289 F. Supp. 2d 1230 (W.D. Wa. 2003) (“*PPA*”), for their argument that, because McGovern was “peer reviewed,” any further scrutiny by this Court (or Defendants’ experts) is precluded. *PPA* does not announce any such rule. On the contrary, *PPA* underscores the importance of scrutinizing key studies relied upon by experts for significant methodological errors.

PPA involved an MDL where the plaintiffs claimed cough and cold and appetite-suppressant products containing phenylpropanolamine (“*PPA*”) caused them to suffer hemorrhagic strokes. The Hemorrhagic Stroke Project (HSP) was the result of research, clinical trials and case studies since 1979, and served as the central epidemiological study in *PPA*. Discussing the defendants’ challenge to expert testimony that relied on the HSP due to claimed flaws, the *PPA* court noted that, “[c]ourts frequently depend on epidemiological studies in determining the reliability of expert testimony.” 289 F. Supp. 2d at 1239.

Further noting that the HSP had grown out of pre-litigation research and was subject to peer review (the main points of apparent interest to Plaintiffs here), the court emphasized the extensive review of the HSP analysis even prior to its submission and ultimate publication in the *New England Journal of Medicine*:

The HSP underwent multiple layers of review. In addition to the FDA and the autonomous SAG, the HSP involved, from its inception, both the NDMA and two defendant-manufacturers. This involvement included approval of the investigators selected, the SAG members, and the study protocol, as well as an opportunity to challenge the study. In fact, in reviewing the study and industry criticisms, the FDA considered many of the same challenges raised here. In rejecting these criticisms, the FDA epidemiologic and statistical reviewers found the study “well designed and executed.” . . .

The epidemiologists found the study’s strengths to include “the clarity of its objectives, the meticulous adherence to sound epidemiology practices in its design and execution, and the consistency of the findings, regardless of the analytical methods.” . . . Indeed, far from finding the study flawed, the FDA’s statistician found the HSP “one of the best planned, conducted and most thoroughly analyzed studies reviewed in the last ten years.”

Id. at 1240.

While observing that scientific studies almost invariably contain flaws, the court concluded that, “[w]hen faced with epidemiological evidence, the court *must* determine whether the flaws compromise the study’s findings.” *Id.* (emphasis added). The court then conducted a “close examination of the arguments and supporting evidence” and found that the HSP’s “flaws,” including any flaws unknown to the FDA and/or the *New England Journal of Medicine*, were either “inaccurately identified [by the defendants] as

flaws or inconsequential to the reliability of the study as a whole. The HSP investigators utilized widely accepted and reliable scientific and epidemiological procedures in conducting the study. Because the court finds the methodology scientifically sound, any flaws that might exist go to the weight afforded the HSP, not its admissibility.” *Id.*

In sum, the *PPA* court did not simply determine that the HSP had been subject to “peer review” and end the inquiry. The court conducted a “close examination of the arguments and supporting evidence” and evaluated the defendants’ challenges to the HSP’s flaws to determine whether the flaws compromised the study’s findings.

B. Unlike the HSP Discussed in *PPA*, McGovern’s Severe Flaws Compromise Its Findings.

The *PPA* court permitted the plaintiffs’ experts to rely on the HSP based on a host of supporting factors, including the way in which the HSP came about, its investigators’ objectivity, the involvement and scrutiny of stakeholders including the FDA, and the pre-study development of a sound and reliable protocol. 289 F. Supp. 2d at 1234–36, 1240. The court concluded that any serious potential flaws had been vetted by independent FDA epidemiological and statistical reviewers, one of whom concluded that the HSP was “one of the best planned, conducted and most thoroughly analyzed studies reviewed in the last ten years.” *Id.* at 1240.

Contrast this with McGovern. McGovern did not arise from independent concerns expressed by the FDA, based on multiple reports from independent medical practitioners over a span of years. Instead, McGovern was launched by a fraud-convicted competitor with an axe to grind. There was no preplanning or careful consideration of the study

design protocol by anyone, let alone a large group of stakeholders with competing interests; rather, the study was “opportunistic,” conducted *ad hoc* after data had been collected and available for analysis. The start date was changed and the surveillance period was selected *after* data had been collected, and the study’s authors conceded serious tabulation errors. McGovern was not subject to thorough and objective vetting by multiple levels of objective oversight. And McGovern was conducted by investigators with clear conflicts of interest—conflicts not disclosed to any reader of the published article, who would not be able to discern a connection between McGovern and Augustine. In the published paper, Albrecht is identified as a “graduate student in statistics” at the University of Minnesota. Indeed, he was a student at the University of Minnesota—part time, and financed by Augustine, while employed by Augustine.¹⁹ Albrecht Dep. 18:12-25.

“*Well-conducted* studies are uniformly admitted.” *PPA*, 289 F. Supp. 2d at 1240 (emphasis added). Determining whether a study is “well-conducted,” however, is not left to the sole province of unnamed and unknown “peer reviewers.” Rather, when “faced with epidemiological evidence, the court *must* determine whether the flaws compromise the study’s findings.” *Id.* (emphasis added). The *PPA* court conducted a “close examination of the argument and supporting evidence” and found the “flaws” asserted by

¹⁹ The “conflict of interest” statement at the end of McGovern vaguely alludes to the existence of some undisclosed relationship: “The author or one or more of the authors have received or will receive benefits for personal or professional use from a commercial party related directly or indirectly to the subject of this article.” McGovern 1544.

the defendants either “inaccurately identified as flaws or inconsequential to the reliability of the study as a whole.” *Id.*

By contrast, the *Viagra* court and others have excluded expert testimony based on “peer-reviewed” studies that had been shown to suffer from major flaws. *See Viagra II*, 658 F. Supp. 2d at 950. Famously, Andrew Wakefield published an article in *The Lancet* in 1998 finding that the measles-mumps-rubella (MMR) vaccine was associated with autism. Unknown to *The Lancet* or even his co-authors, Wakefield received funding from plaintiffs’ litigation counsel preparing to sue vaccine manufacturers. DX32, Godlee et al., “Wakefield’s article linking MMR vaccine and autism was fraudulent,” *BMJ* (Jan. 6, 2011). Ultimately, *The Lancet* retracted the article due, in part, to Wakefield manipulating his raw data to tell a plaintiff-friendly story. DX33, Editors of the Lancet, “Retraction – Ileal-lymphoidnodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children,” 375 *Lancet* 445 (2010). However, ***even prior to retraction***, serious questions about the underlying data led the U.S. Court of Federal Claims to disregard the testimony of experts who relied on the “peer-reviewed” article. *See Cedillo v. Sec’y of Health & Human Servs.*, 2009 WL 331968, at *79, 109–11 (Fed. Cl. Feb. 12, 2009).

Similarly, a New Jersey court excluded an expert relying on his own peer-reviewed article for several flaws made evident by the revelation of the study’s raw data—or lack thereof. *See Palazzolo v. Hoffman La Roche, Inc.*, 2010 WL 363834 (N.J. App. Div. 2010). In that case, the expert had published a peer-reviewed article associating Accutane with depression. *Id.* at *1. After defendants had a chance to review

the study's underlying data, the expert "was repeatedly confronted with problems in the . . . study, including missing data, inaccurate data, and deviations from the methodology he claimed to have followed." *Id.* at *2. The trial court therefore excluded the study, and an appellate court affirmed. *Id.* at *4–5 ("An expert's scientific peers cannot fairly judge the expert's written work, including whether it is worthy of publication, if his article does not accurately represent either the underlying data or what the author did to produce his results.").

This case is like *Viagra II*, *Cedillo*, and *Palazzolo*. Holford and Borak's thorough examination of McGovern, taken together with other evidence adduced during discovery, demonstrate that McGovern's flaws are deep, consequential, incapable of remediation, and so fundamental as to vitiate reliability.

Moreover, McGovern is not "pre-litigation" in the sense Plaintiffs imply. As detailed in Defendants' motion to exclude Plaintiffs' engineering experts, McGovern was part of a suite of studies concocted, conceived, orchestrated, and implemented by Scott Augustine as part of his plan to destroy the Bair Hugger system and increase the sales of his competitor product, the HotDog, by generating and publishing phony science and collaborating with Plaintiffs' lawyers to bring litigation based on that phony science. McGovern is the crown jewel to emerge from what Augustine employee (and McGovern co-author) Mark Albrecht aptly described as the "publication factory." DX31, 7/9/2010 Email from Albrecht to Reed at 1.

IV. HOLFORD AND BORAK MEET THE REQUIREMENTS OF *FRYE-MACK*.

In addition to meeting the Federal Rules' threshold for admissibility, Holford's and Borak's rebuttal opinions are admissible under Minnesota Rule of Evidence 702 and the *Frye-Mack* standard. *See Goeb v. Tharaldson*, 615 N.W.2d 800, 814 (Minn. 2000). Biostatistics and epidemiology are generally accepted fields of scientific study and practice that involve analyzing and re-analyzing biostatistical data. Holford's and Borak's methods are no different than standard practice in those fields. Borak builds upon Holford's standard biostatistical analysis by performing a standard epidemiological causation analysis, in direct rebuttal of Samet's epidemiological causation analysis. Minnesota courts agree that "[t]he function of rebuttal testimony is to explain, repel, counteract or disprove evidence of the adverse party." *Signature Flight Support Corp. v. Cty. of Ramsey*, No. 62-CV-14-3089, 2017 WL 1377751 (Minn. Tax Apr. 7, 2017) (quoting *Aviva Sports, Inc.*, 829 F. Supp. 2d at 834)); *accord Whitney v. Buttrick*, 376 N.W.2d 274, 278 (Minn. Ct. App. 1985) (granting new trial based on district court's improper exclusion of rebuttal expert testimony). It is entirely appropriate, helpful, and relevant for Holford and Borak to critique Plaintiffs' experts, and they should be permitted to do so at trial in Ramsey County as well as in the MDL.

CONCLUSION

Holford's thorough, painstaking, and objective analysis and explication of the biostatistical flaws is methodologically sound, consistent with the highest standards of biostatistical practice, and would be immensely helpful to the jury in understanding the

data manipulation that allowed McGovern to find an “association” between the Bair Hugger system and increased surgical site infections as compared to the HotDog. Holford’s opinion is confined to the factual record and limited to his biostatistical analysis. This is uniquely within his area of expertise as a biostatistician. *See Kadas*, 255 F.3d at 362–63 (“The question whether a study is responsible and therefore admissible under the *Daubert* standard is different from the weight to be accorded the significance of a particular correlation found by the study. It is for the Judge to say, *on the basis of the evidence of a trained statistician*, whether a particular significance level, in the context of a particular study in a particular case, is too low to make the study worth the consideration of the judge or jury.” (Emphasis added)). Building on Holford, Borak applies his epidemiological and medical expertise and his review of the literature to conclude that Plaintiffs’ medical causation experts’ conclusions are not supported by scientifically convincing evidence. Holford and Borak both meet the legal requirements of Federal Rules of Evidence 702 and 703, *Daubert*, and Minnesota law. The Court should deny Plaintiffs’ motions.

Dated: October 3, 2017

Respectfully submitted,

s/Corey L. Gordon

Jerry W. Blackwell

MN Atty ID No. 0186867

Corey L. Gordon

MN Atty ID No. 125726

Benjamin W. Hulse

MN Atty ID No. 0390952

**Attorneys for Defendants 3M Company
and Arizant Healthcare Inc.**

BLACKWELL BURKE P.A.

431 South Seventh Street, Suite 2500

Minneapolis, MN 55415

T: (612) 343-3200 F: (612) 343-3205

blackwell@blackwellburke.com

cgordon@blackwellburke.com

bhulse@blackwellburke.com

Bridget M. Ahmann

MN Atty ID No. 016611x

**Attorney for Defendants 3M Company
and Arizant Healthcare Inc.**

FAEGRE BAKER DANIELS LLP

2200 Wells Fargo Center

90 South Seventh Street

Minneapolis, MN 55402

T: (612) 766-7000 F: (612) 766-1600

bridget.ahmann@faegrebd.com

ADDENDUM

Responses to Selected Misstatements by Plaintiffs' of Borak's Report and Testimony

Plaintiff's Statement	Borak's Response
<p><u>Page 1:</u> “Dr. Borak claims that the McGovern study has no validity whatsoever”</p>	<p>Borak concluded that McGovern lacked <i>internal</i> validity and therefore lacks external validity. Borak Rpt. ¶ 59.</p>
<p><u>Page 1:</u> “Dr. Borak did not independently review the findings of the McGovern study.”</p> <p><u>Page 7:</u> “Dr. Borak did not independently review any of those documents in detail.”</p>	<p>He <i>did</i> review the McGovern study, the related reports and depositions, and the data tables that were exhibits to the Albrecht (Ex. 10) and McGovern (Ex 16) depositions. He did not independently perform statistical reanalyses; he relied on Prof. Holford's statistical reanalysis. Borak Dep. 38:4–39:10.</p>
<p><u>Page 2:</u> “He refused to consider non-epidemiological sources of causation evidence.”</p>	<p>Borak considered Plaintiffs' cited non-epidemiological, “mechanistic” studies and concludes: “In the absence of valid evidence of a causal association between BH and SSI, it can only be said that the mechanistic studies are coherent with a hypothetical increase in SSI. Hypothetical associations are not sufficient to sustain an inference of causation.” Borak Rpt. ¶ 71.</p>
<p><u>Page 6:</u> “Dr. Borak suggests that the McGovern study provides <i>no</i> evidence of an association....”</p>	<p>Borak concludes that McGovern is <i>insufficient</i> evidence of an association:</p> <p>Q. Your opinion is that the use of the Bair Hugger is not associated with an increased risk of infection.</p> <p>A. I believe there is insufficient evidence to make that statement. I believe that it has been associated in two problematic studies. I don't know that there is sufficient evidence otherwise. Borak Dep. 14:8-14.</p>
<p><u>Page 9:</u></p>	<p>Reed confirmed the Gillson paper's report and that</p>

“Dr. Borak misleadingly states that Dr. Reed acknowledged that Wansbeck was a ‘high outlier.’”	Northumbria made “loads of changes” in response. Reed Dep. 66:1–68:2. Regarding infection reduction efforts, Gillson says that the English NHS identified Northumbria as an “outlier,” and Reed acknowledged that “we were identified . . . as having a high infection rate.” <i>Id.</i> at 66:7–8; 69:12–13.
Page 9: “Dr. Borak also ignored 3M’s own analysis of national infection rates.”	Borak did not review an internal 3M document listing infection rates because the rates included <i>all</i> infections, including superficial ones. Another internal document from the same deposition shows that the deep joint infection rate is typically less than 1%. DX34, PowerPoint presentation “Forced-air warming systems, OR air quality, and orthopedic SSI,” at 3MBH01185207.
Page 11: “Dr. Borak next speculates that change in antibiotic regimens from gentamicin to gentamicin plus teicoplanin . . . confounded the findings . . . yet he cites no studies.”	Borak was not “speculating.” He relied specifically on Reed’s statements: “Reed reported that ‘our infection rate doubled when we went to Gentamicin’ (18) and that following introduction of prophylactic gentamicin, the rate of return to theater because of SSI increased significantly, from 0.66% to 1.52% ($p < 0.01$): ‘We recommend that single dose Gentamicin (4.5 mg/kg) alone is not used as prophylaxis for joint replacement’ (25).” Borak Rpt. ¶ 34.
Page 12: “Dr. Borak’s second theory . . . Dr. Reed’s statement, as cited by Dr. Borak, does not apply to DJI but wound infections in general.”	That is inaccurate. The Reed comments were specifically focused on “deep infection[s].” See DX35, A. Brister, “Infection Control in Orthopaedic Surgery,” <i>Clin. Services. J.</i> 2(Nov. 9, 2011).
Pages 13–14: “Dr. Borak inappropriately concluded that the antithrombotic regimens were a ‘highly significant confounder’ of the McGovern study.”	Plaintiffs misstate Borak’s conclusion. He concludes that when the Jensen study was reanalyzed using the McGovern surveillance period, the use of antithrombotics was statistically significantly associated with deep joint infections. By analogy, the antithrombotic regimen would have been a confounder in the McGovern study. Borak Rpt. ¶ 44.

<p><u>Pages 13–14:</u> “The Jensen study corroborates the subsequent findings of the Jameson and Reed study.”</p>	<p>The Jameson report does not inform the question regarding post-op infections. Borak Rpt. ¶ 45.</p>
<p><u>Page 15</u> “Drs. Borak and Holford ignore these contrary studies [the RECORD studies] despite their coherence.”</p>	<p>Borak did not ignore contrary studies. He considered them and found that they did not inform the question. Borak Dep. 206:9–208:14.</p>
<p><u>Page 15:</u> “Inexplicably, however, he [Borak] omitted a series of studies – ranging from the Jensen and Jameson papers all the way to the four well known RECORD trials.”</p>	<p>Borak did not omit those papers. He discussed the Jensen and Jameson papers in both his report and deposition. Borak Rpt. ¶ 45; Borak Dep. 155:1–5. He discussed the RECORD trials during his deposition and explained their lack of relevance. Borak Dep. 206:24–208:14.</p>
<p><u>Page 20:</u> “The most bizarre opinion Dr. Borak espouses is that a psychological phenomenon dubbed the Hawthorne Effect confounded the McGovern study. . . . He admits he has ‘no idea’ if any of the hospital staff knew they were participating in the McGovern study because the data was gathered after the time period.”</p>	<p>The Hawthorne Effect occurs when people change their behaviors because they are aware they are being observed. Contrary to Plaintiffs’ assertion, it does not depend on whether they know that they are going to be part of a later study. Borak Rpt. ¶ 48 (citing Eckmanns et al., “Compliance with antiseptic hand rub use in intensive care units: the Hawthorne effect,” 27 <i>Infection Control & Hosp. Epidemiology</i> 931 (2006). It is likely that OR staff behavior at Wansbeck Hospital changed because of the introduction of numerous infection reduction measures, the close observation of OR staff actions, and the introduction of a “strict theatre etiquette that is adhered to by all.” Gillson 73. If those changes led to improved infection rates (for example, improved hand washing or less talking-and-moving by staff during surgery), and if one warming device was used before and the other after the implementation of the infection reduction measures and “strict theatre etiquette,” then the resulting behavior changes would have confounded comparisons between the warming devices. Borak Rpt. ¶ 48.</p>